



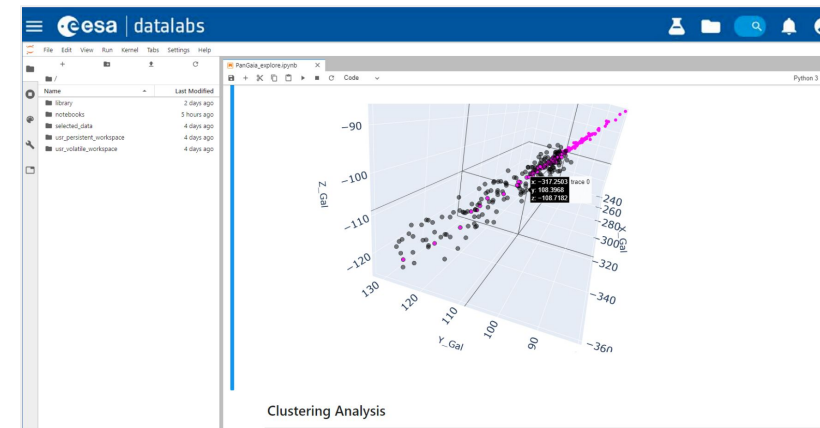
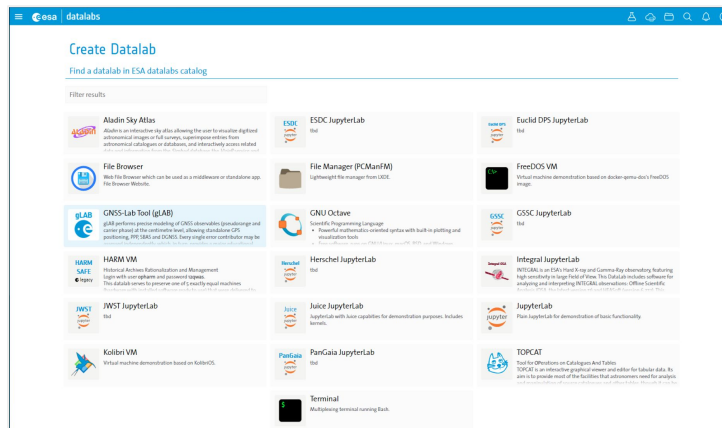
Rosetta: a container-centric science platform for interactive, resource intensive data analysis

Stefano Alberto Russo - INAF OATS



What is a science platform?

“A science platform is an environment designed to offer users a smoother experience when interacting with computing and storage resources”



What is a science platform?

“A science platform is an environment designed to offer users a smoother experience when interacting with computing and storage resources”

- We usually assume a web-based environment built on top of Jupyter notebooks or similar software
- Other approaches or definitions may be also feasible (e.g. full interactive desktop access).



Example definition

"a web Portal, designed to provide essential data access and visualization services through a simple-to-use website, a Notebook environment, that will provide a Jupyter Notebook-like interface, based on JupyterLab, enabling next-to-the-data analysis"

Juric et al. 2017



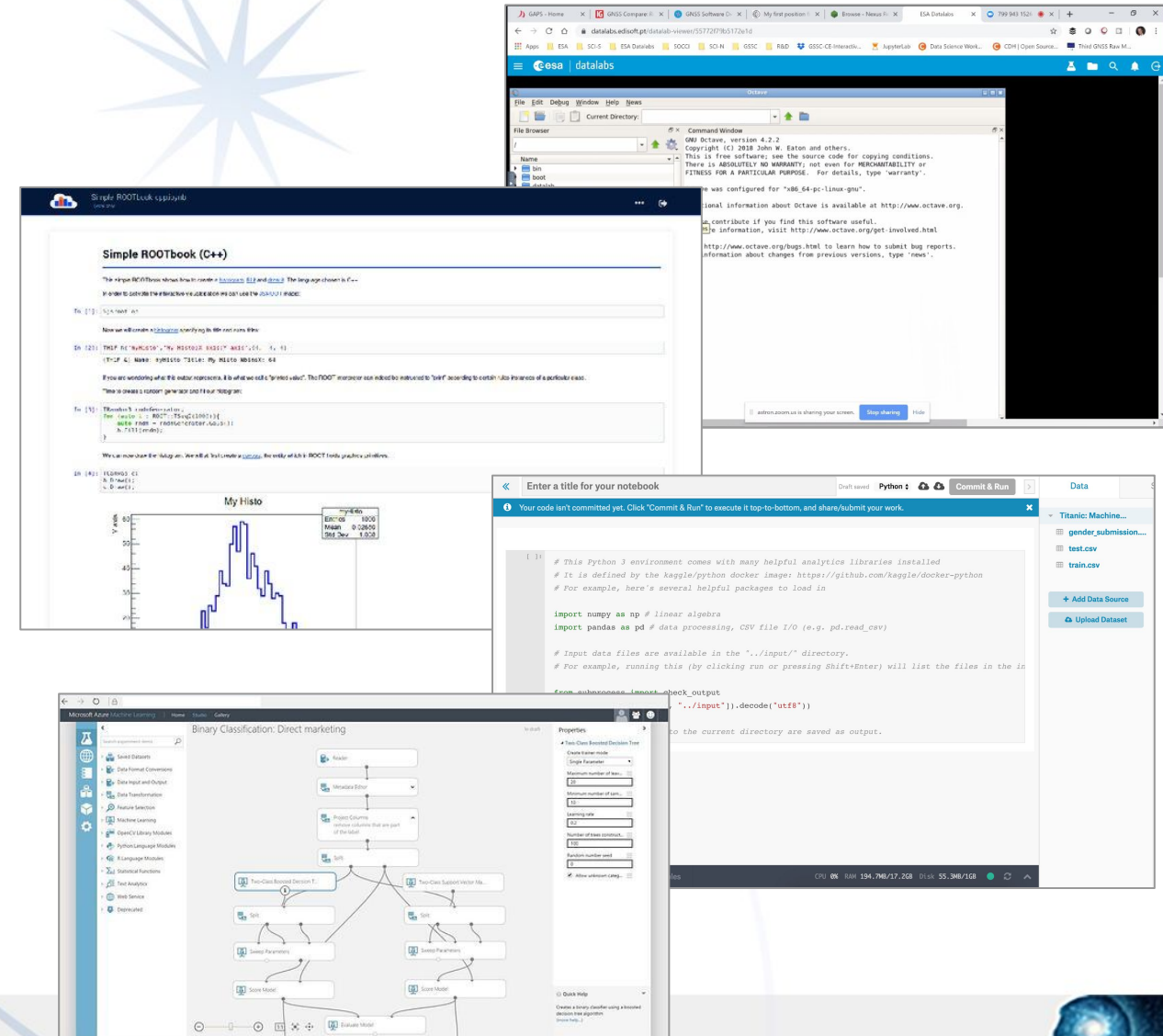
Some examples

Public Research/ Academia:

- CERN SWAN
- ESA Datalabs
- LSST Science Platform
- ...

Private Research / Industry:

- Google Colabs
- Kaggle Kernels
- Azure Machine Learning Studio
- ...



Limitations

There are some strong limitations with the current science platforms:

1. A user cannot run in an environment not supported by the platform
→ from a different Python version to a different Linux distribution.
→ this is a reproducibility issue!!
2. A user is tight to a specific interface, with little options for other ones
→ a web-based notebook (Jupyter or similar) interface makes it impossible to run native GUI applications (common in Astrophysics).
3. Poor HPC support, hard to integrate
→ the focus is usually more on making the data accessible rather than to run resource-intensive analysis

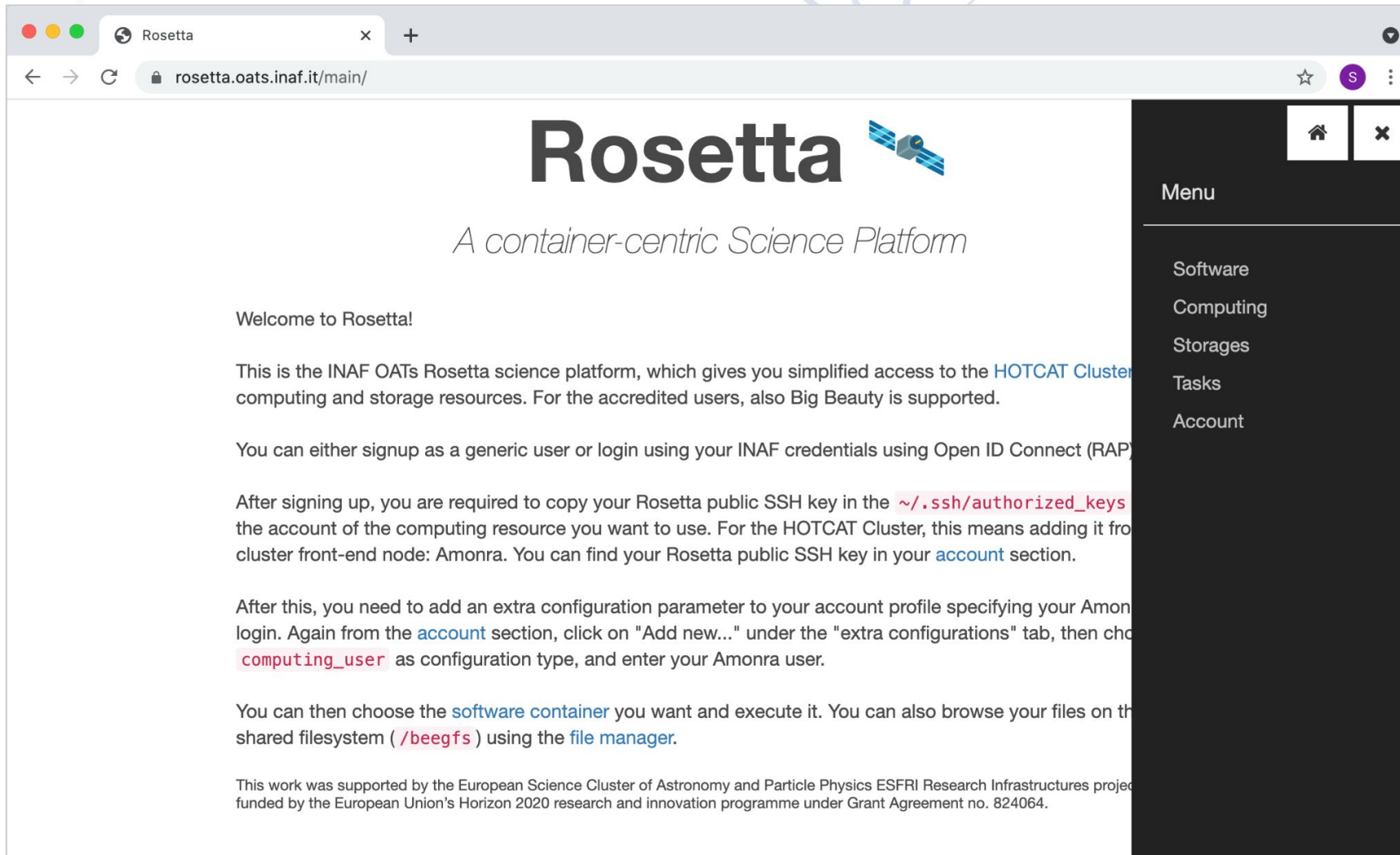



Introducing Rosetta

- A web-based science platform supporting both pre-defined and custom analysis environments
- Executes user tasks in software containers on:
 - local CPUs;
 - standalone computing resources
 - HPC clusters and Cloud systems
- Microservice-oriented approach: each task runs in its own container and expose its interface directly to the user:
 - *Users can choose from different software and interfaces*
 - *Users can add their own containers*



Rosetta main page



Rosetta 

A container-centric Science Platform

Welcome to Rosetta!

This is the INAF OATs Rosetta science platform, which gives you simplified access to the [HOTCAT Cluster](#) computing and storage resources. For the accredited users, also Big Beauty is supported.

You can either signup as a generic user or login using your INAF credentials using Open ID Connect (RAP).

After signing up, you are required to copy your Rosetta public SSH key in the `~/.ssh/authorized_keys` the account of the computing resource you want to use. For the HOTCAT Cluster, this means adding it from cluster front-end node: Amonra. You can find your Rosetta public SSH key in your [account](#) section.

After this, you need to add an extra configuration parameter to your account profile specifying your Amonra login. Again from the [account](#) section, click on "Add new..." under the "extra configurations" tab, then choose `computing_user` as configuration type, and enter your Amonra user.

You can then choose the [software container](#) you want and execute it. You can also browse your files on the shared filesystem (`/beegfs`) using the [file manager](#).

This work was supported by the European Science Cluster of Astronomy and Particle Physics ESFRI Research Infrastructures project funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 824064.

Menu

- Software
- Computing
- Storages
- Tasks
- Account



The software “catalogue”

Rosetta

rosetta.oats.inaf.it/containers/

Software Containers

Search... All Go

Basic Desktop

A basic desktop environment. Provides a terminal, a file manager, a web browser and other generic applications.

Image: `exact/swc/basicdesktop`

Tag: `v0.3.0`

CASADesktop v5.6.1-8

CASA, the Common Astronomy Software Applications package, is the primary data processing software for the Atacama Lar...

Image: `exact/swc/casadesktop`

Tag: `v5.6.1-8`

Jupyter Data Science Lab

The official Jupyter Lab. The Data Science variant, which includes libraries for data analysis from the Julia, Python, and R...

Image: `jupyter/scipy-notebook`

Tag: `lab-3.1.17`

Jupyter Lab

The official Jupyter Lab. The Scipy variant, which includes popular packages from the scientific Python ecosystem.

Image: `jupyter/scipy-notebook`

Tag: `lab-3.2.2`

Jupyter Notebook

A Jupyter Notebook server with basic Python numerical data analysis libraries as Pandas and Numpy.

Image: `exact/swc/jupyternotebook`

Tag: `v0.3.0`

Lofar_visualization

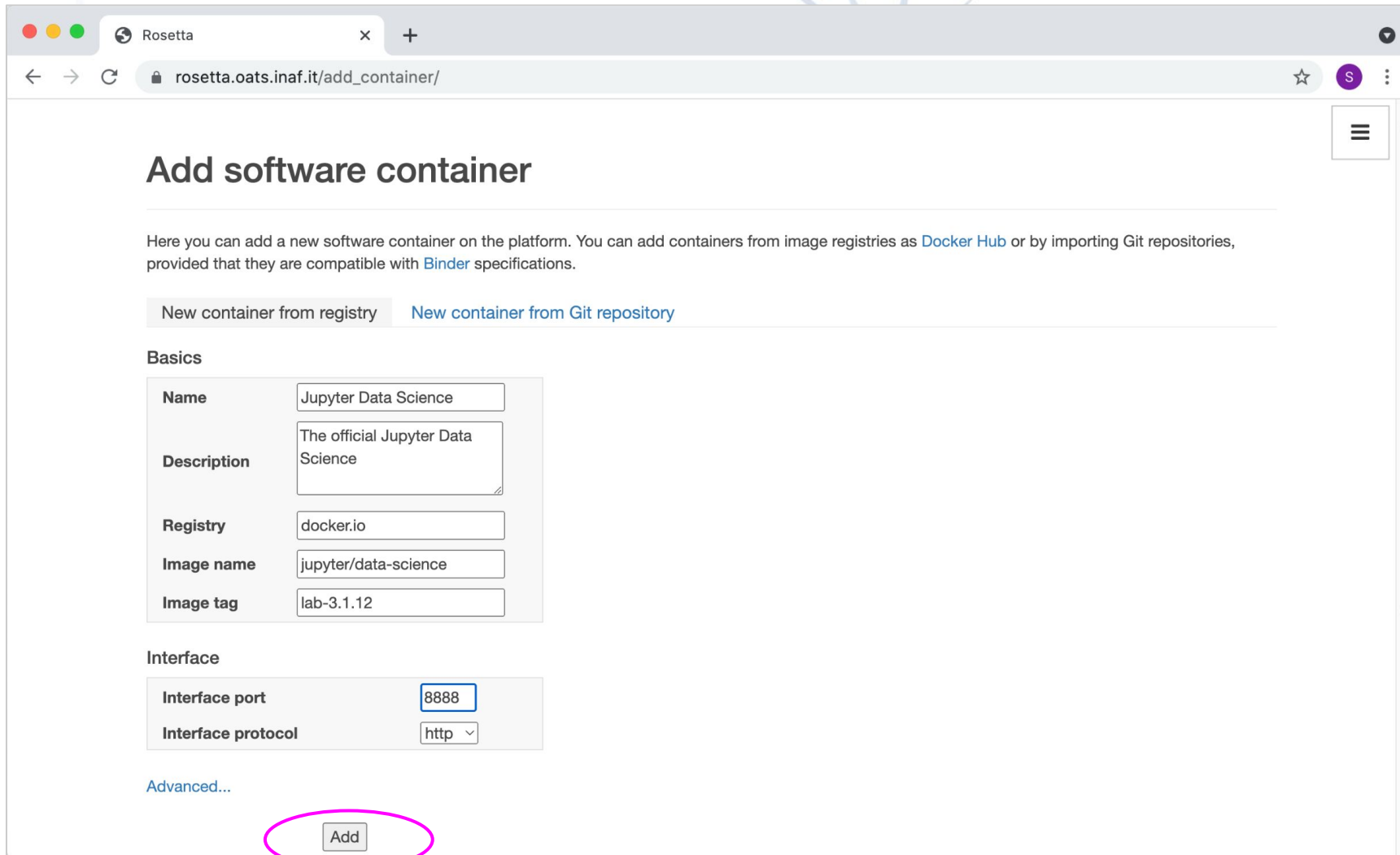
The LOFAR (Low Frequency Array) data visualization toolkit.

Image: `exact/swc/lofar_visualization`

Tag: `latest`



Adding new software



The screenshot shows a web browser window with the address bar displaying "rosetta.oats.inaf.it/add_container/". The page title is "Add software container". Below the title, there is a paragraph explaining that users can add containers from image registries like Docker Hub or by importing Git repositories, provided they are compatible with Binder specifications. There are two tabs: "New container from registry" (selected) and "New container from Git repository". Under the "Basics" section, there are five input fields: "Name" (Jupyter Data Science), "Description" (The official Jupyter Data Science), "Registry" (docker.io), "Image name" (jupyter/data-science), and "Image tag" (lab-3.1.12). Under the "Interface" section, there are two fields: "Interface port" (8888) and "Interface protocol" (http). At the bottom, there is a link "Advanced..." and a button "Add" which is circled in pink.

Add software container

Here you can add a new software container on the platform. You can add containers from image registries as [Docker Hub](#) or by importing Git repositories, provided that they are compatible with [Binder](#) specifications.

[New container from registry](#) [New container from Git repository](#)

Basics

Name	Jupyter Data Science
Description	The official Jupyter Data Science
Registry	docker.io
Image name	jupyter/data-science
Image tag	lab-3.1.12

Interface

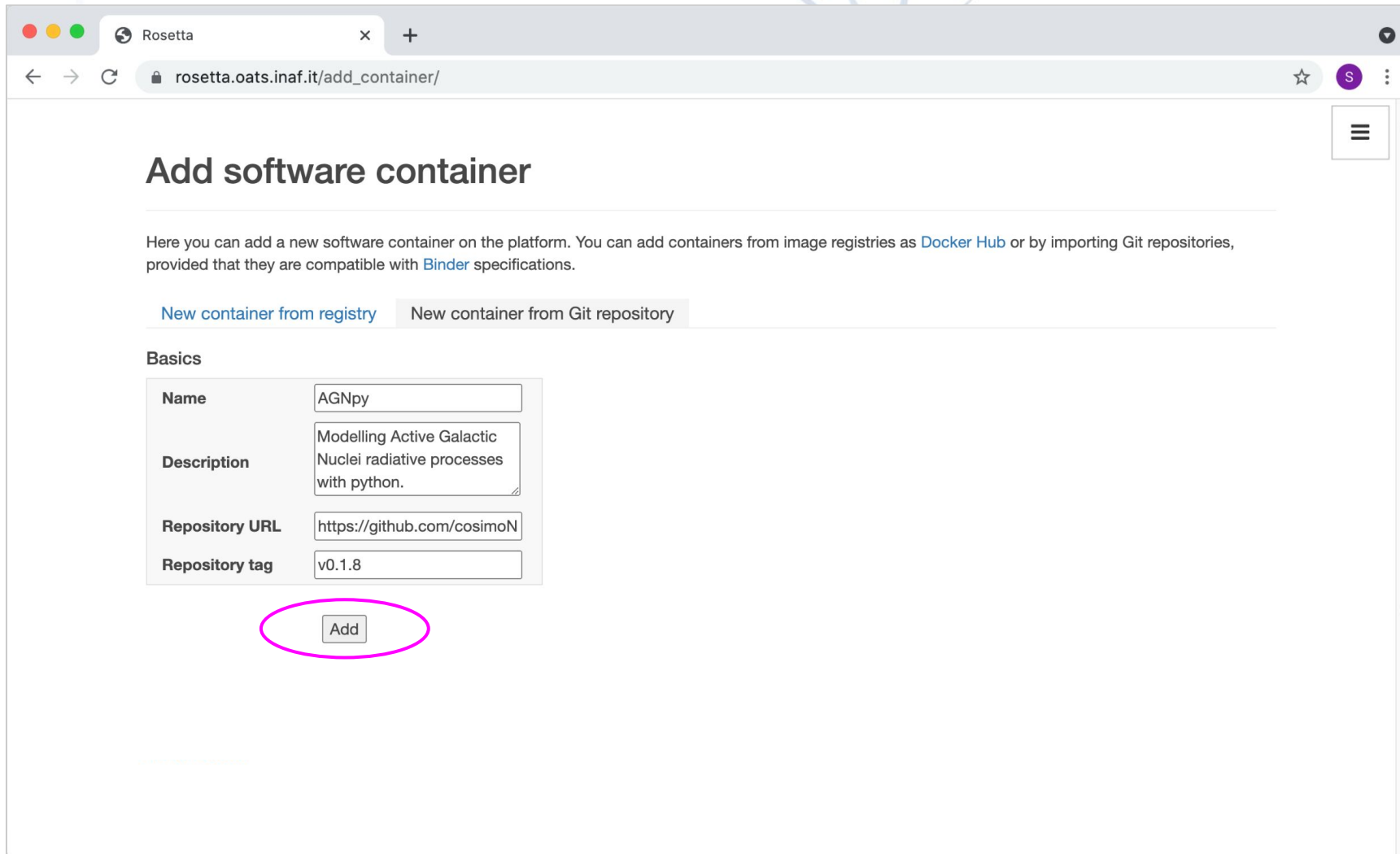
Interface port	8888
Interface protocol	http

[Advanced...](#)

[Add](#)



Adding new software



The screenshot shows a web browser window with the address bar displaying "rosetta.oats.inaf.it/add_container/". The page title is "Add software container". Below the title, there is a paragraph explaining that users can add containers from image registries like Docker Hub or by importing Git repositories, provided they are compatible with Binder specifications. There are two tabs: "New container from registry" (selected) and "New container from Git repository". Under the "Basics" section, there is a form with four fields: "Name" (AGNpy), "Description" (Modelling Active Galactic Nuclei radiative processes with python.), "Repository URL" (https://github.com/cosimoN), and "Repository tag" (v0.1.8). The "Add" button is circled in pink.

Add software container

Here you can add a new software container on the platform. You can add containers from image registries as [Docker Hub](#) or by importing Git repositories, provided that they are compatible with [Binder](#) specifications.

[New container from registry](#) [New container from Git repository](#)

Basics

Name	<input type="text" value="AGNpy"/>
Description	<input type="text" value="Modelling Active Galactic Nuclei radiative processes with python."/>
Repository URL	<input type="text" value="https://github.com/cosimoN"/>
Repository tag	<input type="text" value="v0.1.8"/>



Creating a new task

Software Containers

Search... All Go

Basic Desktop

A basic desktop environment. Provides a terminal, a file manager, a web browser and other generic applications.

Image: `exact/swc/basicdesktop`

Tag: `v0.3.0`

▶

CASADesktop v5.6.1-8

CASA, the Common Astronomy Software Applications package, is the primary data processing software for the Atacama Lar...

Image: `exact/swc/casadesktop`

Tag: `v5.6.1-8`

▶

Jupyter Data Science Lab

The official Jupyter Lab. The Data Science variant, which includes libraries for data analysis from the Julia, Python, and R...

Image: `jupyter/scipy-notebook`

Tag: `lab-3.1.17`

▶

Jupyter Lab

The official Jupyter Lab. The Scipy variant, which includes popular packages from the scientific Python ecosystem.

Image: `jupyter/scipy-notebook`

Tag: `lab-3.2.2`

▶

Jupyter Notebook

A Jupyter Notebook server with basic Python numerical data analysis libraries as Pandas and Numpy.

Image: `exact/swc/jupyternotebook`

Tag: `v0.3.0`

▶

Lofar_visualization

The LOFAR (Low Frequency Array) data visualization toolkit.

Image: `exact/swc/lofar_visualization`

Tag: `latest`

▶



Creating a new task

Rosetta

rosetta.oats.inaf.it/create_task/

☆ S

New Task

Software container

[CASADesktop v5.6.1-8](#)

CASA, the Common Astronomy Software Applications package.
Image: `exact/swc/casadesktop`
Tag: `v5.6.1-8`

Computing resource

[Amonra](#)

A powerful multi-core computing resource.
Type: standalone
Storages:
Home @ `/storages/home`

Details and confirm

Task name

My CASA Task

Auth token

590ddec8-4229-403a-b38e-1a339489467b

A randomly generated token to be used as task password. Usually automatically handled by Rosetta when login-in to the task.

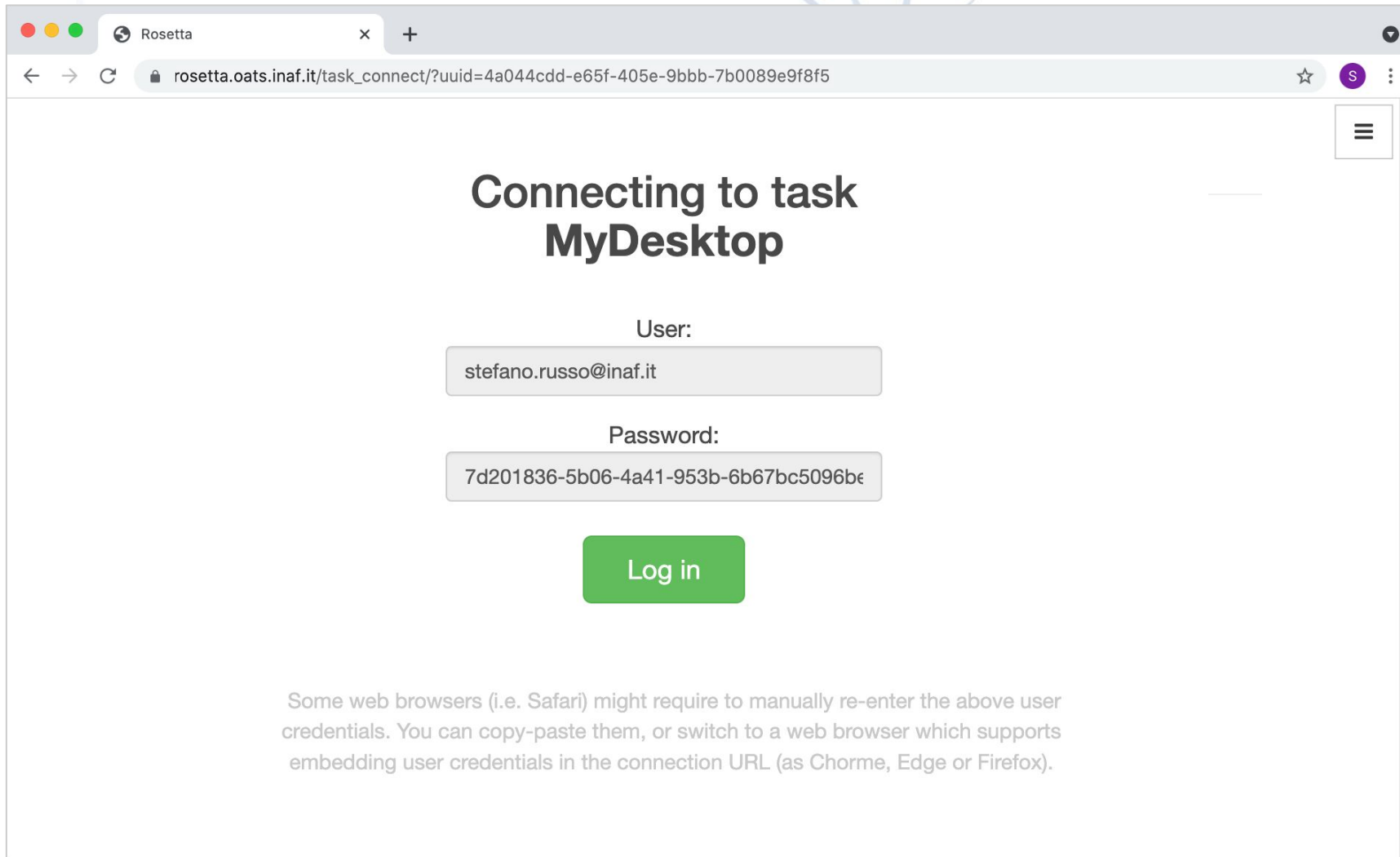
I understand that files saved or modified in the task, if not explicitly saved to a persistent storage, will be LOST upon task completion.

☒

Create task



Creating a new task



The screenshot shows a web browser window with the title 'Rosetta'. The address bar displays the URL 'rosetta.oats.inaf.it/task_connect/?uuid=4a044cdd-e65f-405e-9bbb-7b0089e9f8f5'. The main content area has the heading 'Connecting to task MyDesktop'. Below this, there are two input fields: 'User:' with the value 'stefano.russo@inaf.it' and 'Password:' with the value '7d201836-5b06-4a41-953b-6b67bc5096be'. A green 'Log in' button is positioned below the password field. At the bottom, a note states: 'Some web browsers (i.e. Safari) might require to manually re-enter the above user credentials. You can copy-paste them, or switch to a web browser which supports embedding user credentials in the connection URL (as Chrome, Edge or Firefox).'

Connecting to task
MyDesktop

User:

stefano.russo@inaf.it

Password:

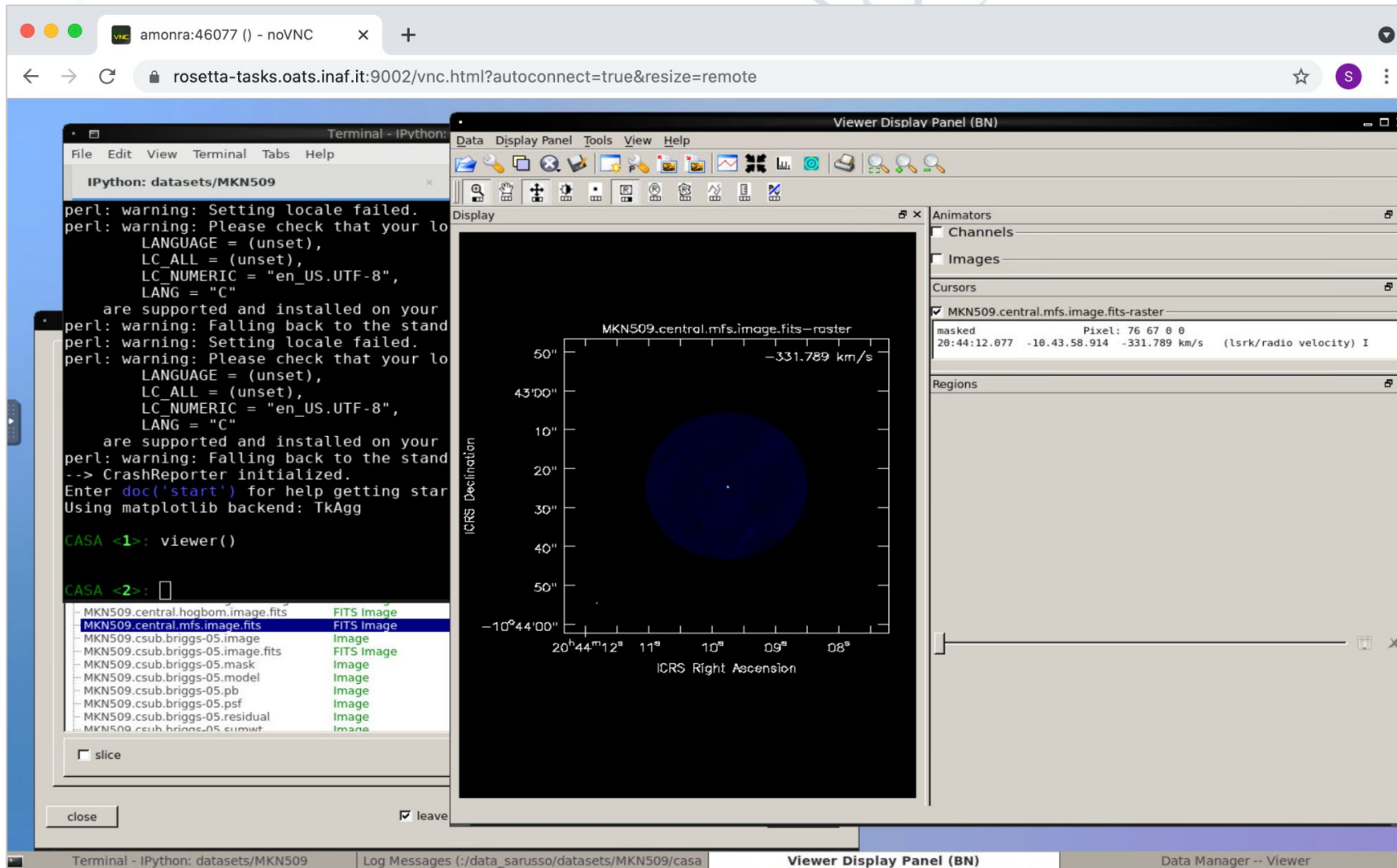
7d201836-5b06-4a41-953b-6b67bc5096be

Log in

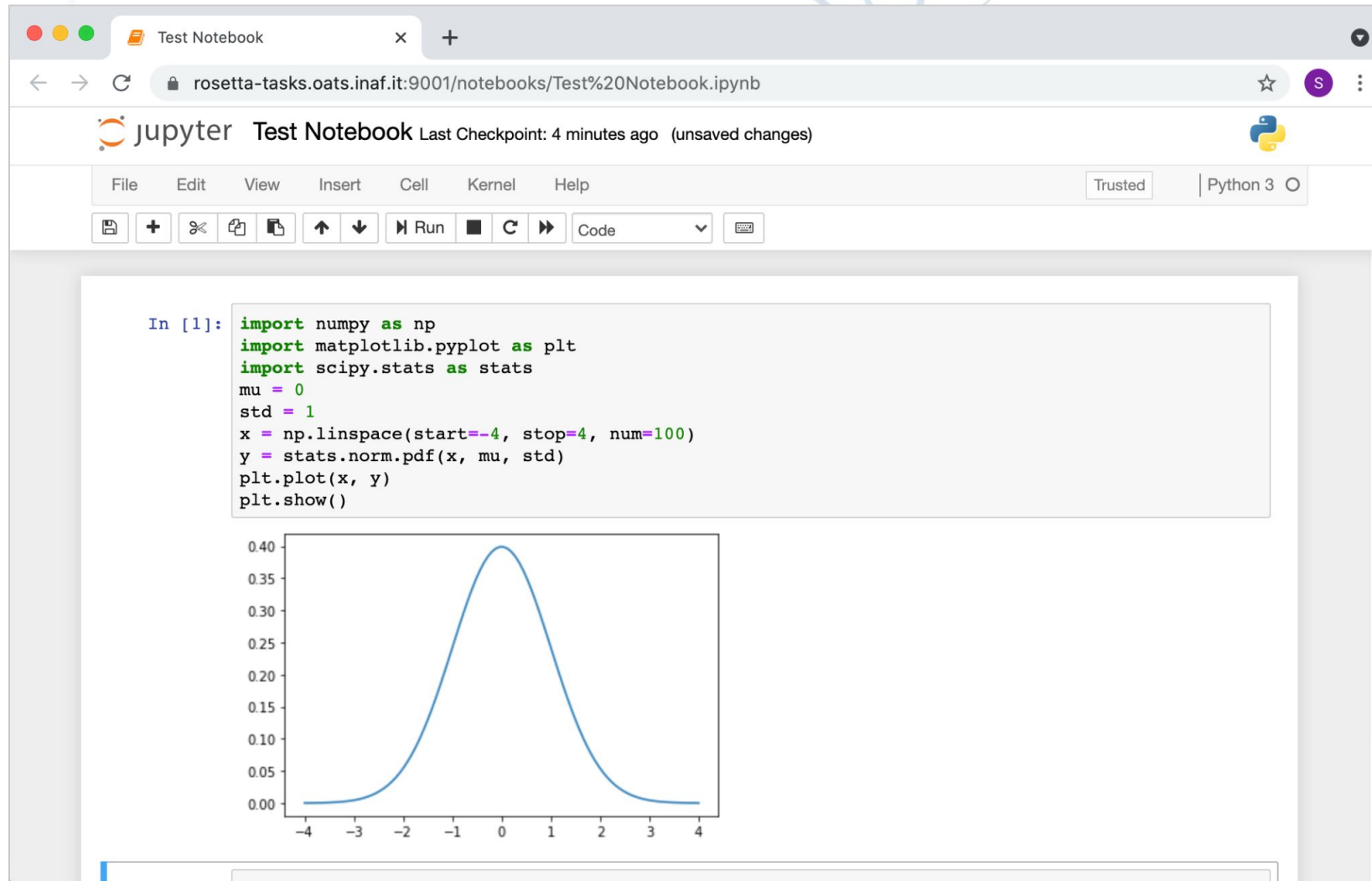
Some web browsers (i.e. Safari) might require to manually re-enter the above user credentials. You can copy-paste them, or switch to a web browser which supports embedding user credentials in the connection URL (as Chrome, Edge or Firefox).



A remote desktop task



A Jupyter Notebook task



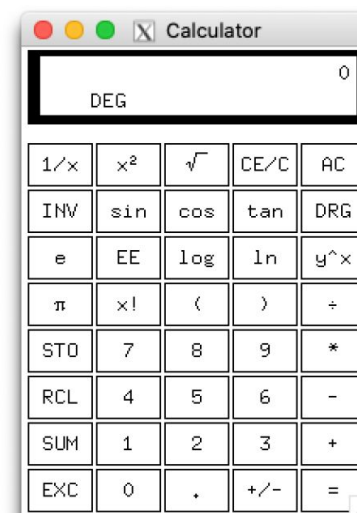
A SSH server task

```

ste@Stes-MacBookAir:~ $ ssh -X -p7000 metauser@rosetta-tasks.oats.inaf.it
metauser@rosetta-tasks.oats.inaf.it's password:
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

metauser@task-0d378dfd:~$ pwd
/home/metauser
metauser@task-0d378dfd:~$ ls -alh
total 44K
drwxr-xr-x 6 metauser metauser 4.0K Nov 17 23:55 .
drwxrwxrwx 1 root      root    4.0K Nov 17 23:53 ..
-rw----- 1 metauser metauser  50 Nov 17 23:55 .Xauthority
-rw----- 1 metauser metauser 123 Nov 17 23:55 .bash_history
-rw-r--r-- 1 metauser metauser 220 Apr  4  2018 .bash_logout
-rw-r--r-- 1 metauser metauser 3.8K Nov 17 23:54 .bashrc
drwx----- 2 metauser metauser 4.0K Nov 17 23:53 .cache
-rw-r--r-- 1 metauser metauser   0 Nov 17 23:53 .initialized
drwxrwxr-x 3 metauser metauser 4.0K Nov 17 23:54 .local
drwxr-xr-x 2 metauser metauser 4.0K Nov  4 19:58 .logs
-rw-r--r-- 1 metauser metauser 807 Apr  4  2018 .profile
drwxr-xr-x 2 metauser metauser 4.0K Nov 17 23:53 custom_ssh
metauser@task-0d378dfd:~$ xcalc

```



The file manager

Rosetta

rosetta.oats.inaf.it/files/

Storages

Home > Amonra:Home > output_data > Pre-Facet-Cal > parsets

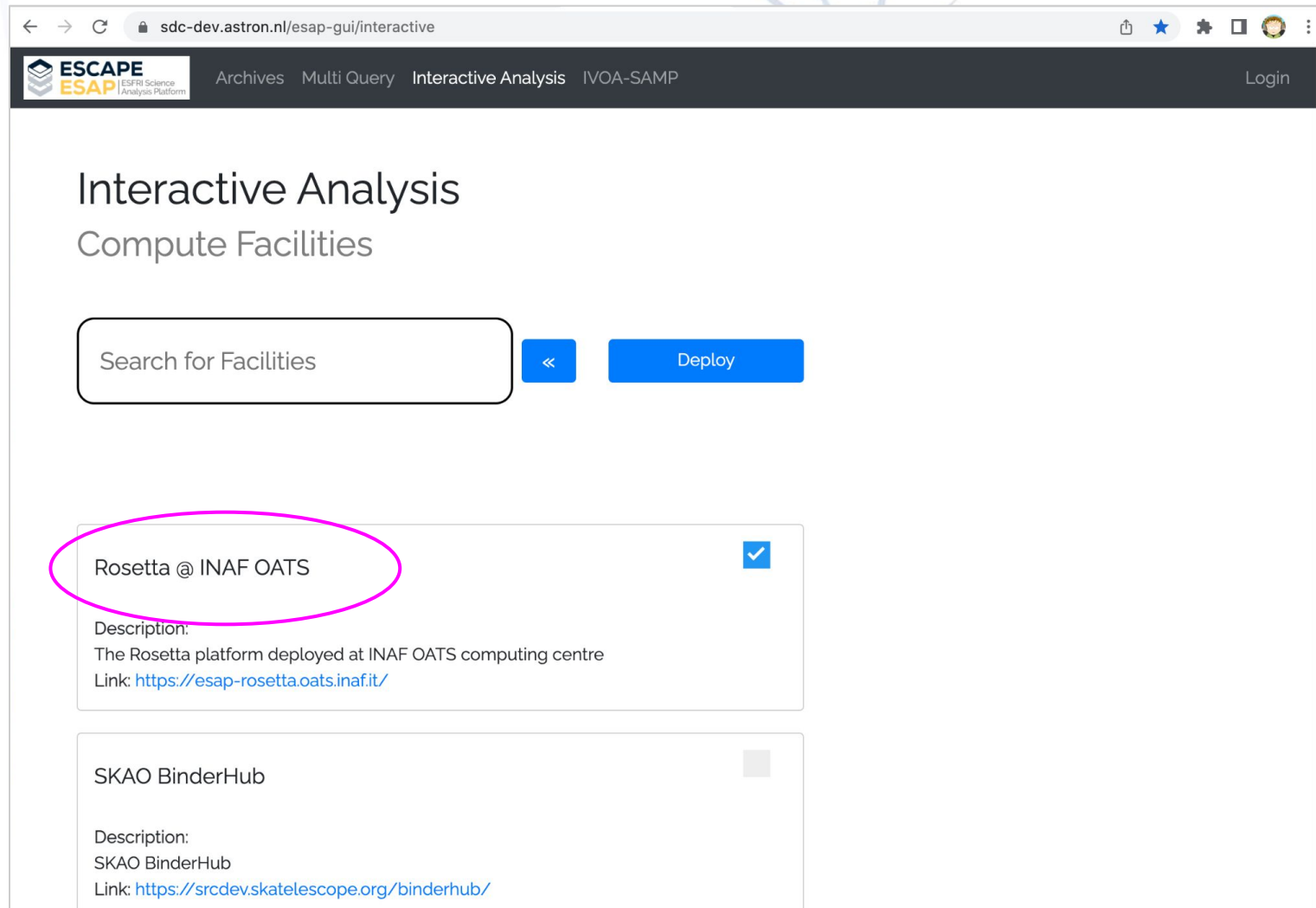
Upload New folder

apply_beam1.parset	parset 305 bytes	Mar 26, 2020, 16:15
apply_FR.parset	parset 296 bytes	Mar 26, 2020, 16:15
apply_FR1.parset	parset 296 bytes	Mar 26, 2020, 16:15
apply_PA.parset	parset 287 bytes	Mar 26, 2020, 16:15
apply_PA1.parset	parset 287 bytes	Mar 26, 2020, 16:15
calib_cal.parset	parset 372 bytes	Mar 26, 2020, 16:15
calib_cal1.parset	parset 372 bytes	Mar 26, 2020, 16:15
calib_cal2.parset	parset 363 bytes	Mar 26, 2020, 16:15
calib_cal21.parset	parset 363 bytes	Mar 26, 2020, 16:15
check_Ateam_separation.parset	parset 115 bytes	Mar 26, 2020, 16:15
h5exp_cal_bandpass.parset	parset 131 bytes	Mar 26, 2020, 16:15
h5exp_cal_FR.parset	parset 130 bytes	Mar 26, 2020, 16:15
h5exp_cal_ion.parset	parset 128 bytes	Mar 26, 2020, 16:15
h5exp_cal_PA.parset	parset 131 bytes	Mar 26, 2020, 16:15
h5imp_cal_bandpass.parset	parset 44 bytes	Mar 26, 2020, 16:15

38 items - Size: 7.06 Kb



ESAP integration



The screenshot shows a web browser window with the URL `sdsc-dev.astron.nl/esap-gui/interactive`. The page header includes the ESAP logo and navigation links: Archives, Multi Query, Interactive Analysis, and IVOA-SAMP. A 'Login' button is in the top right. The main content area is titled 'Interactive Analysis' and 'Compute Facilities'. It features a search bar with the text 'Search for Facilities', a '<<' button, and a 'Deploy' button. Below this, there are two facility cards. The first card, 'Rosetta @ INAF OATS', is highlighted with a red oval and has a checked checkbox. Its description is 'The Rosetta platform deployed at INAF OATS computing centre' and its link is <https://esap-rosetta.oats.inaf.it/>. The second card, 'SKAO BinderHub', has an unchecked checkbox. Its description is 'SKAO BinderHub' and its link is <https://srcdev.skatelescope.org/binderhub/>.

sdsc-dev.astron.nl/esap-gui/interactive

ESCAPE ESAP ESFRI Science Analysis Platform

Archives Multi Query Interactive Analysis IVOA-SAMP Login

Interactive Analysis

Compute Facilities

Search for Facilities << Deploy

Rosetta @ INAF OATS ☒

Description:
The Rosetta platform deployed at INAF OATS computing centre
Link: <https://esap-rosetta.oats.inaf.it/>

SKAO BinderHub ☐

Description:
SKAO BinderHub
Link: <https://srcdev.skatelescope.org/binderhub/>



Rosetta architecture

Rosetta architecture is based on two main ingredients:

- 1) Microservices
- 2) Software containers



Rosetta architecture: microservices

Microservices are independent and self-contained units that perform a given and well-defined task, using a well defined *interface*.

→ From just summing two numbers to running a neural network.

Microservices are completely decoupled from the underlying infrastructure and from each other. Encapsulation is maximum.

In Rosetta, each user task is framed as a microservice



Rosetta architecture: microservices

Examples of interfaces:

- REST APIs
- HTTP
- SSH
- RPC
- etc.



Rosetta architecture: containers

Software containers are lightweight, standalone, executable packages of software that includes everything needed to run an application:

→ code, runtime, system tools, system libraries and settings.

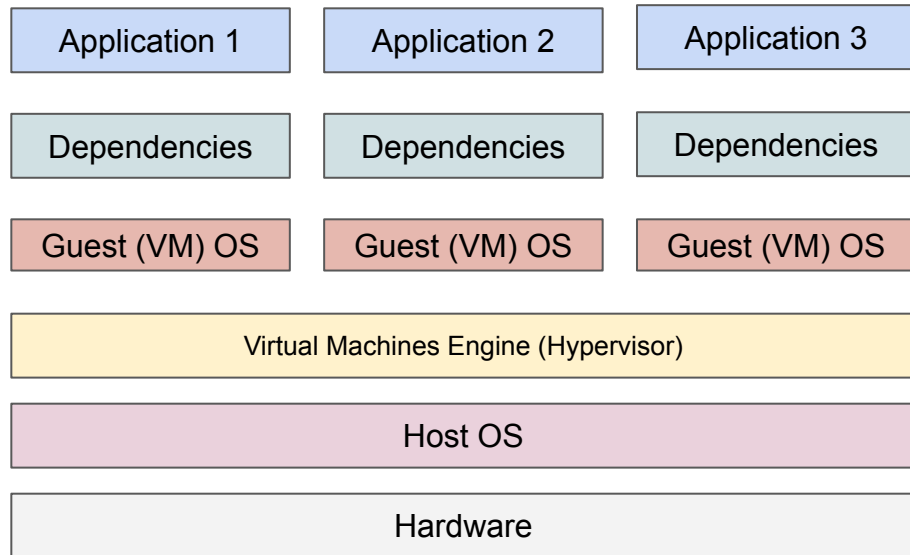
They are a solution to the problem of how to get software to run reliably when moved from one computing environment to another.

Rosetta uses them to wrap the user task microservices

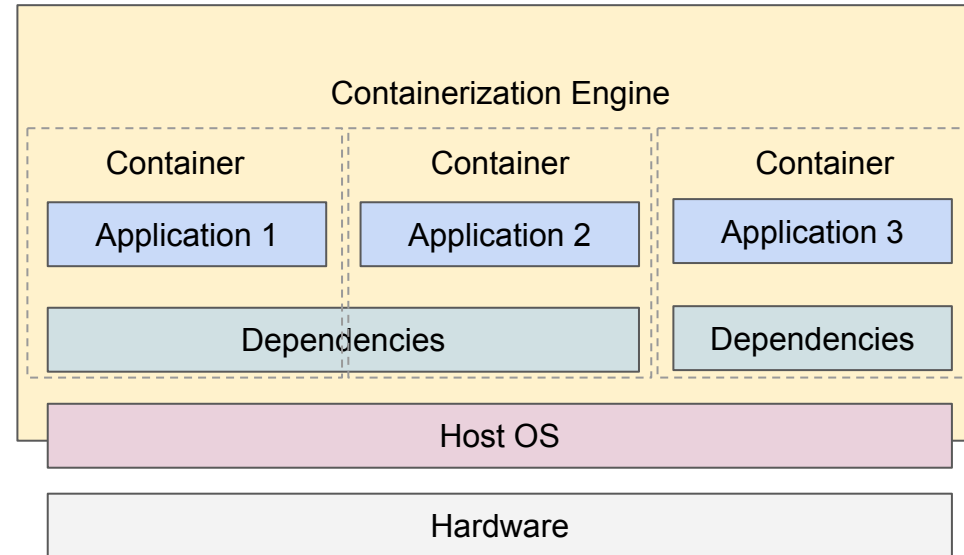


Rosetta architecture: containers

Virtual Machines

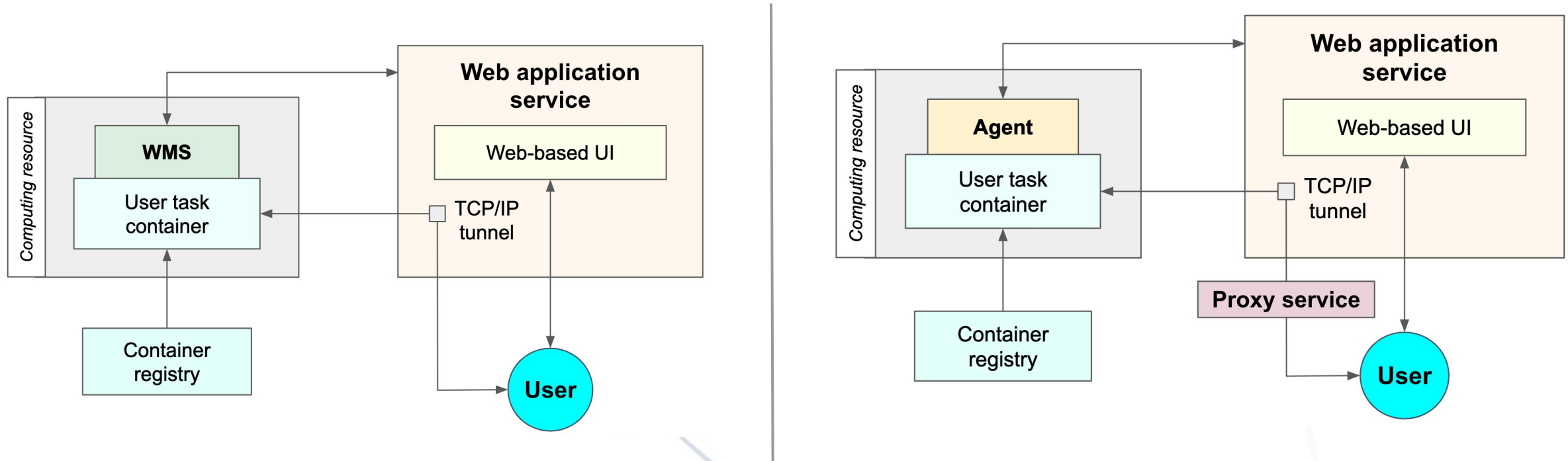


Containers

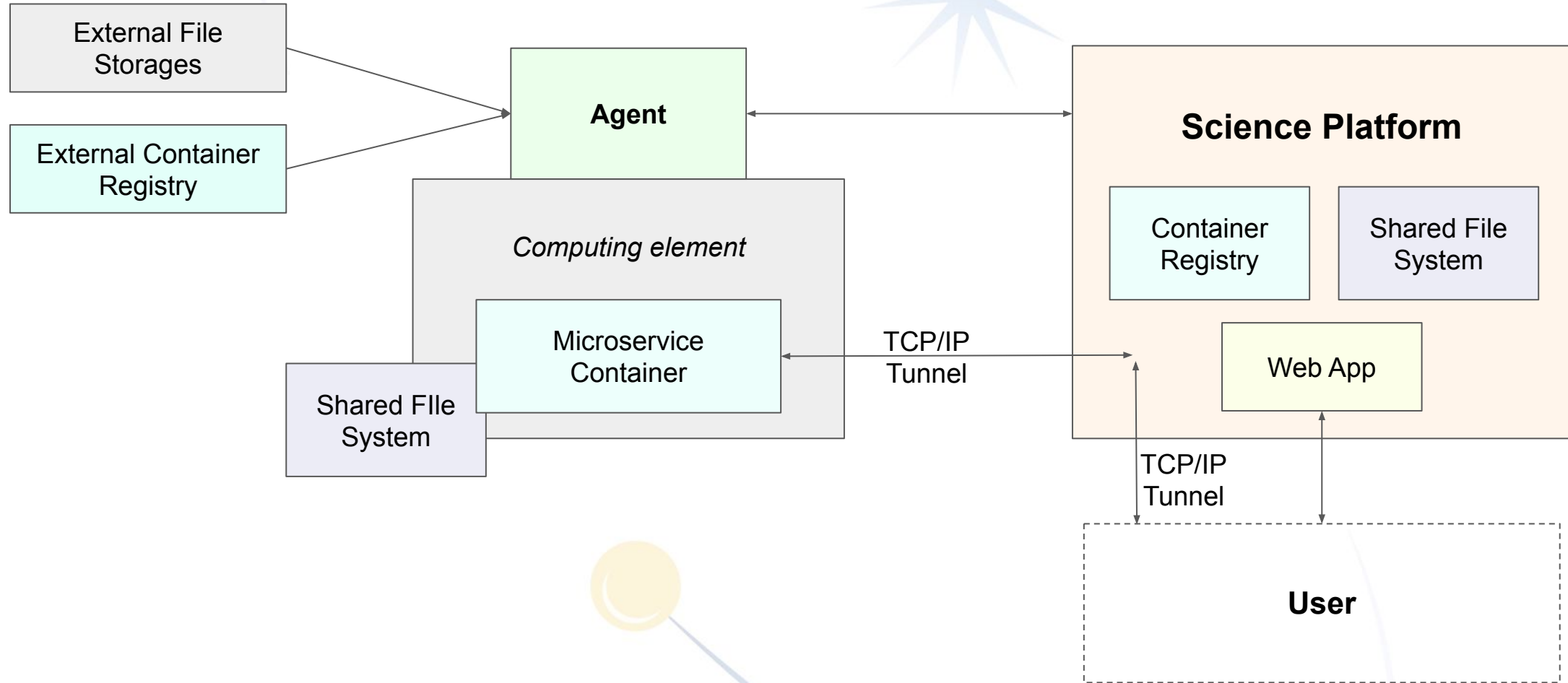


Rosetta architecture: task orchestration

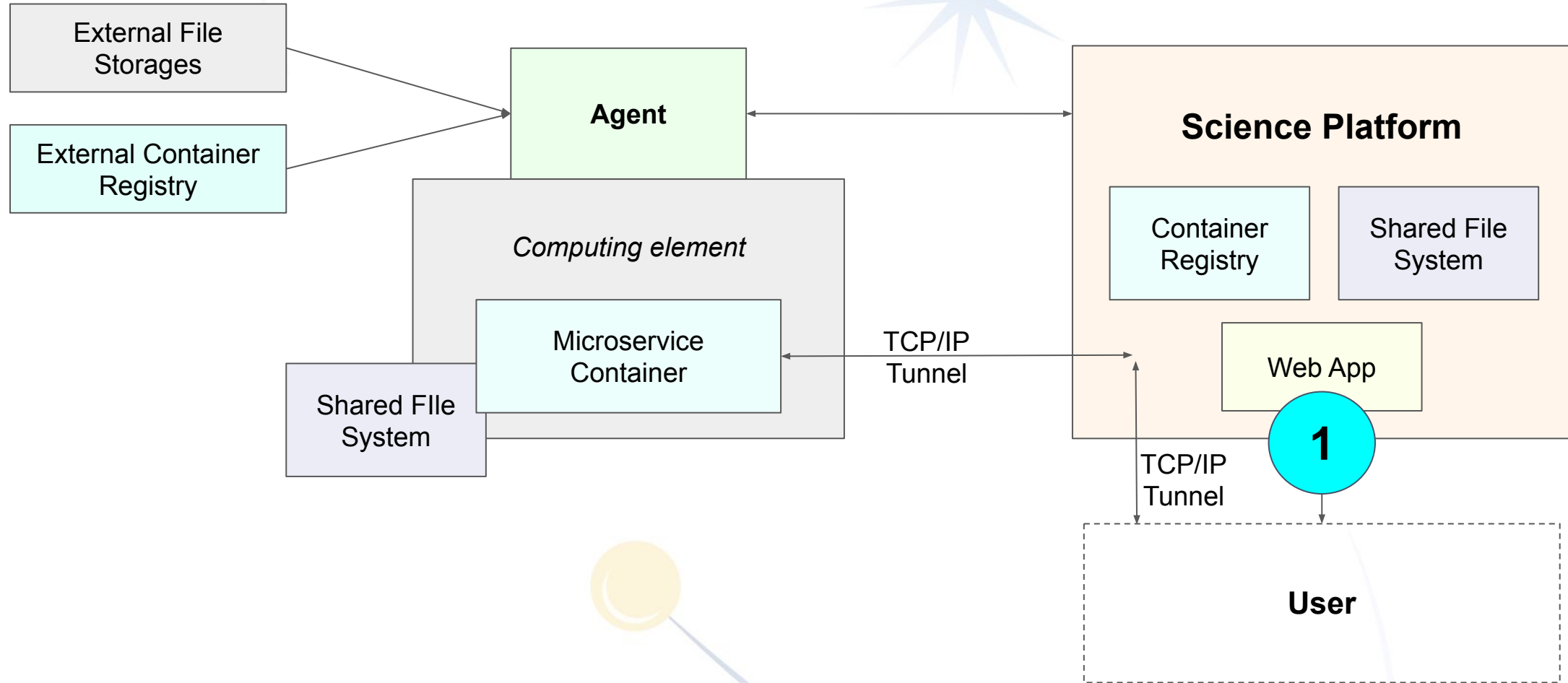
Task orchestration include sending the task for execution and setting up the all the tunneling required to reach its interfaces. It can rely on the WMS or use an *agent*.



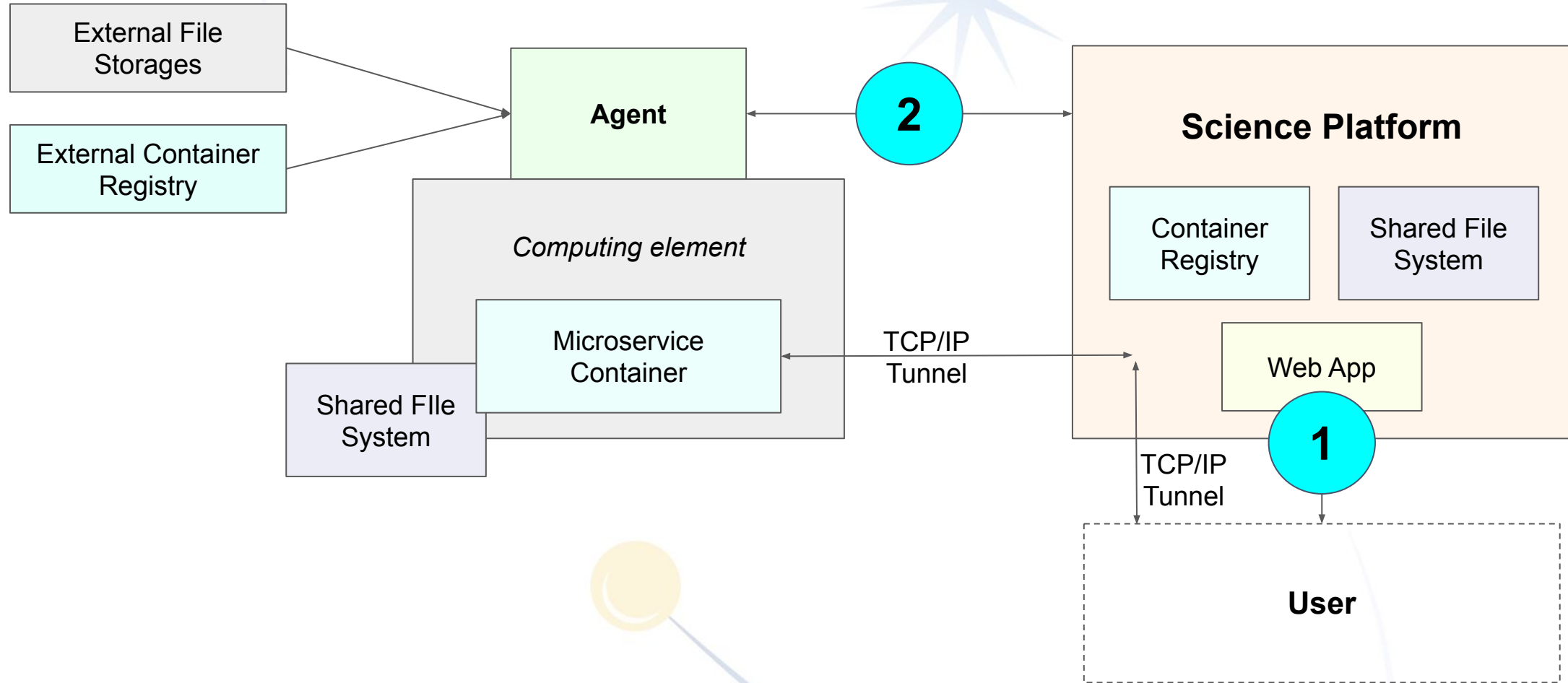
Rosetta architecture: task orchestration



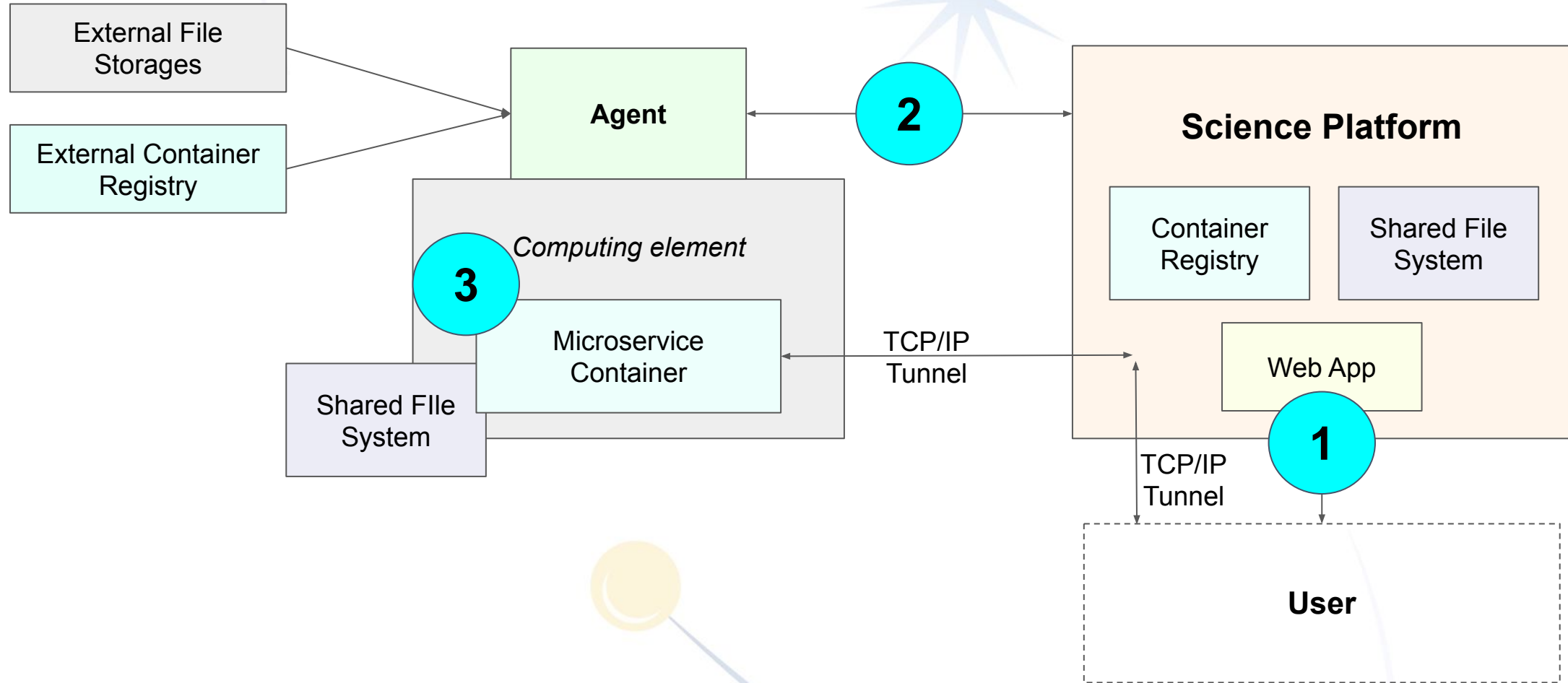
Rosetta architecture: task orchestration



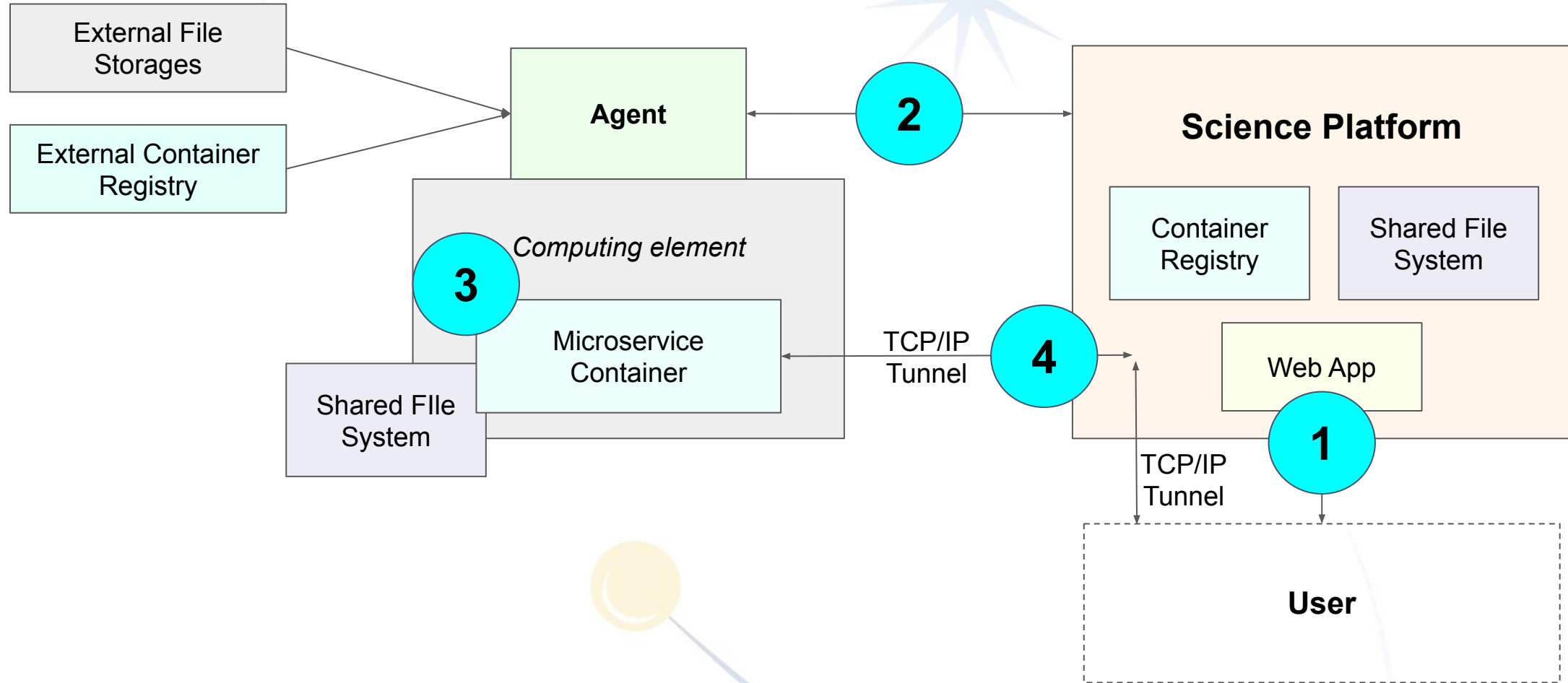
Rosetta architecture: task orchestration



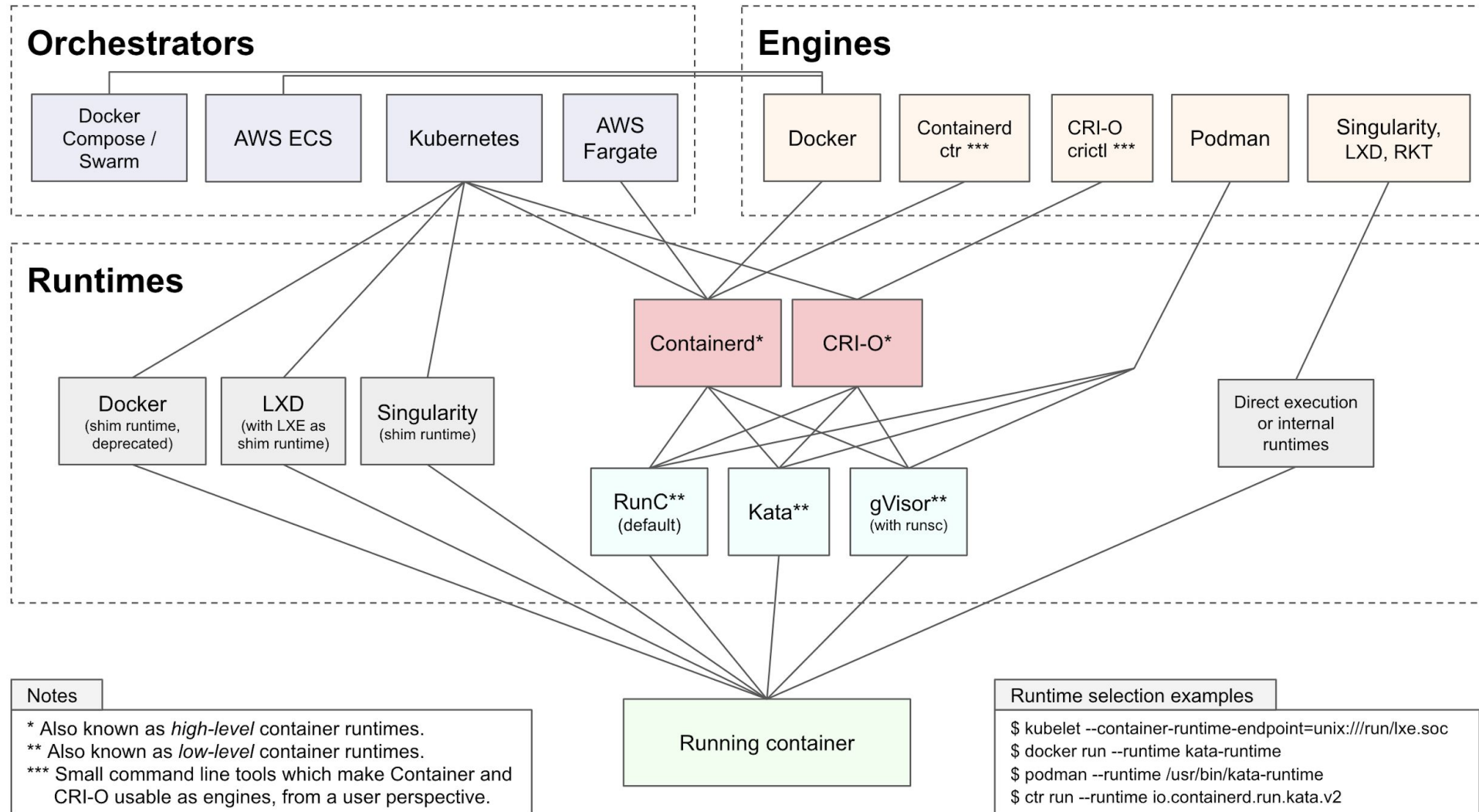
Rosetta architecture: task orchestration



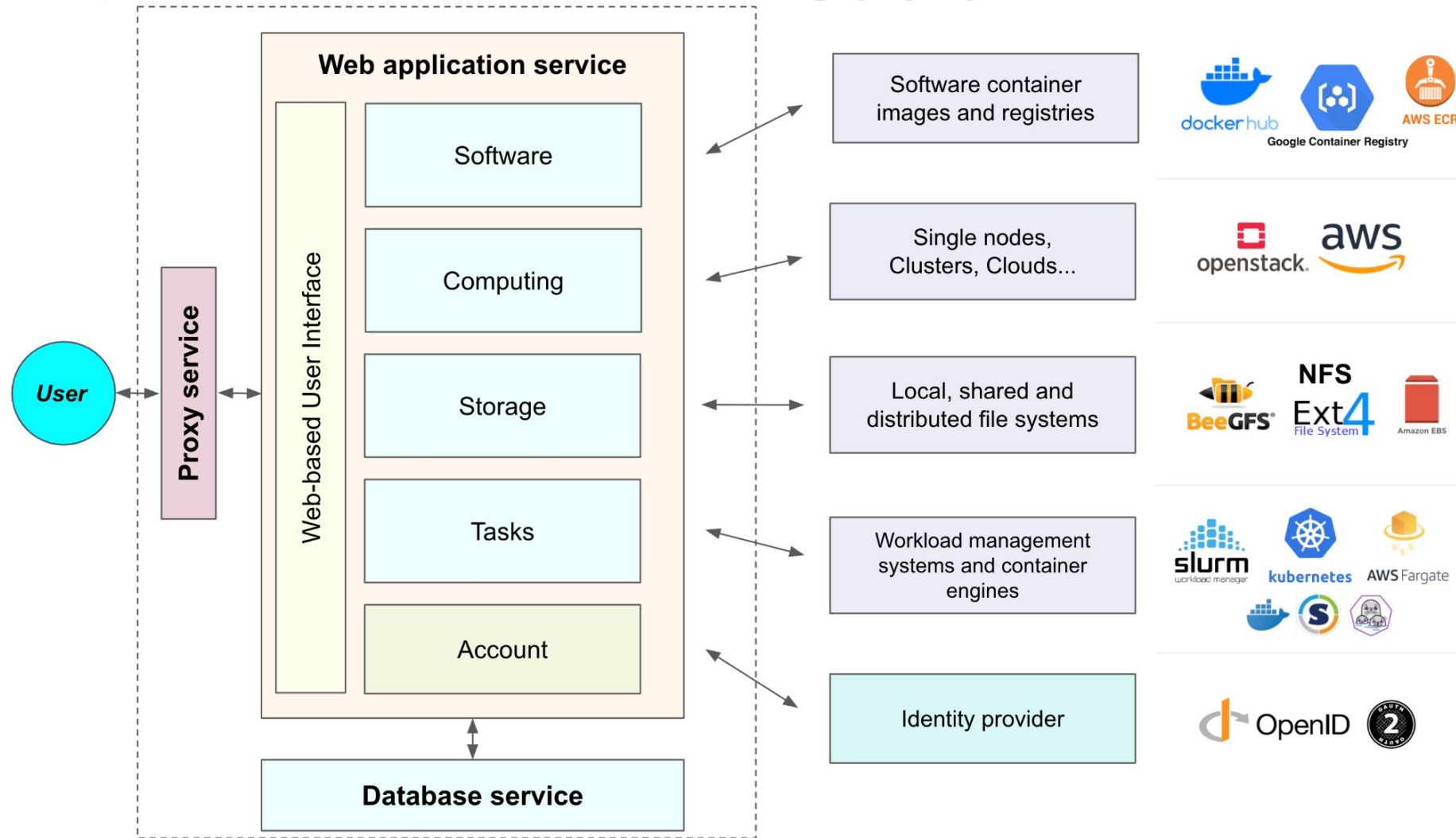
Rosetta architecture: task orchestration



Parenthesis: containerization landscape



Rosetta: an umbrella for n technologies



Wrapping up

- **Rosetta** is a **Science Platform** which allows to run:
 - **well established software** (*pre-loaded on the platform*)
 - **custom software** and/or dependencies (*added by the users*)
→ *both as: Jupyter Notebooks, GUI applications, command line tools etc.*
- Supports **HPC, Cloud** and **data intensive** systems
 - easy to integrate with existing WMS
- **Containers** together with **microservice**-level isolation make it **safe** to let users add their own software and provide a **robust architecture**
- Huge improvements in terms of **reproducibility**



The logo graphic for ESCAPE features a stylized blue starburst at the top, a thin blue orbital line curving around it, and a small yellow circle at the bottom left of the orbit.

ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

THANKS!

Stefano Alberto Russo - INAF OATS

